Enhanced Ocean Prediction using Ensemble Kalman Filter Techniques

Dr Peter R. Oke
CSIRO Marine and Atmospheric Research and Wealth from Oceans Flagship Program
GPO Box 1538 Hobart TAS 7001 Australia
phone: (+61 3) 6232-5387   fax: (+61 3) 6232-5123   email: peter.oke@csiro.au

LONG-TERM GOALS

To apply optimal data assimilation techniques to ocean circulation models in order to improve short-range prediction of mesoscale circulation.

OBJECTIVES

The immediate scientific objective of this research project is to develop a data assimilation system, based on ensemble Kalman filter (EnKF) techniques, and to apply this system to a realistic eddy-resolving ocean circulation model.

APPROACH

The basic elements of an ensemble-based data assimilation system include a system for collating and preparing observations, combining observations with a model now-cast, initializing and running an ensemble of forecasts and estimating model and observation errors. Each of these components are currently under development and are being systematically tested on a suite of models ranging from a simple linear model (Evensen 2004), to a small highly non-linear model (Lorenz and Emmanuel 1998), and finally to an idealised and realistic configuration of an ocean general circulation model (MOM4.0; Griffies et al. 2004).

Under this project, we have explored the rationale for different ensemble-based assimilation algorithms and techniques; and compared the performance of different filters for a suite of small models (Oke et al. 2006; Sakov et al. 2006). Subsequently, we have developed a new formulation for the EnKF that we refer to as the deterministic EnKF (DEnKF; Sakov and Oke 2006).

We have begun to assess some of the above-mentioned developments through their application to an ocean general circulation model. The assimilation system we have used for this is the Bluelink Ocean Data Assimilation System (BODAS) that has been developed by the P.I. under a closely related project called BLUElink (see Related Projects for details). BODAS is a modular ensemble-based data assimilation system that computes an analysis of the three-dimensional ocean circulation by combining the relevant elements of an ensemble of anomalies of model variables, a model forecast and a range of observations. To date, an ensemble optimal interpolation (EnOI; Oke et al. 2002) version of BODAS has been applied to the Ocean Forecasting Australia Model (OFAM; Schiller et al. 2006), a global

Report Documentation Page

1. REPORT DATE **30 SEP 2006**	2. REPORT TYPE	3. DATES COVERED **00-00-2006 to 00-00-2006**

4. TITLE AND SUBTITLE **Enhanced Ocean Prediction using Ensemble Kalman Filter Techniques**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Commonwealth Scientific and Industrial Research Organisation (CSIRO),Marine and Atmospheric Research and Wealth from Oceans Flagship Program,GPO Box 1538,Hobart TAS 7001 Australia,**	8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release; distribution unlimited**

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**11**	

configuration of MOM4.0, with eddy-resolving resolution around Australia. BODAS and OFAM have been combined and implemented into the Ocean Model, Analysis and Prediction System (OceanMAPS) that is currently undergoing operational trials at the Australian Bureau of Meteorology. Under this project we have developed an EnKF version of BODAS. This system is currently being evaluated.

Key individuals:

Dr. Peter Oke is the P.I. on this project and leads the data assimilation group at CSIRO Marine and Atmospheric Research (CMAR). Dr. Oke is responsible for the development of BODAS under the BLUElink project.

Dr. Pavel Sakov is currently investigating algorithms for ensemble data assimilation and methods for optimal array design under a related project.

Dr. Stuart Corney was hired by CMAR in April 2004 to conduct research under this project. Dr Corney has developed a prototype EnKF system based on BODAS and has applied it to a regional configuration of MOM4.0.

The US-based researchers involved in this project are Dr. H. Ngodock, Dr. G. Jacobs and Dr. R. Miller. All have a long history of research in the area of ocean data assimilation.

WORK COMPLETED

We have completed a study that compared the performance of an EnKF with perturbed observations (Burgers et al. 1998) and EnOI when applied to a two-variable linear system that is similar to that of Evensen (2004). In this study we have investigated the impact of localisation (Houtekamer and Mitchel 2001; Hamill et al. 2001) on a filter's performance and on the dynamical balances that are inherent in the model (Oke et al. 2006).

We have explored the theoretical basis for different formulations of the ensemble square-root filters (ESRF; e.g., Tippett et al. 2003) and shown that only a sub-set of these algorithms work, in practice, as they are intended to work in theory. Additionally, through a series of numerical experiments we have compared the performance of different filters when applied to two small models (Sakov et al. 2006).

We have proposed a new formulation for the EnKF that combines the simplicity and flexibility of the traditional EnKF with the robustness and superior performance of the ESRF (Sakov and Oke 2006a). We call this formulation the DEnKF.

A prototype EnKF system, based on BODAS, has been developed and applied to an idealised configuration of MOM4.0 that models the mesoscale ocean circulation in Bass Strait (between Victoria and Tasmania, Australia). This prototype system is currently being evaluated.

RESULTS

Impacts of localisation in EnOI and EnKF:

The performance of an inexpensive, EnOI scheme that uses a stationary ensemble of model anomalies to approximate forecast error covariances, is compared with that of an EnKF with perturbed observations. The model to which the methods are applied is a pair of "perfect" one-dimensional, linear advection equations for two related variables. While EnOI is sub-optimal, it can give results that are comparable to those of the EnKF (Figure 1). The computational cost of EnOI is typically about 1/N times that of the EnKF, where N is the ensemble size. We suggest that EnOI may provide a practical and cost-effective alternative to the EnKF for some applications where computational cost is a limiting factor. We demonstrate that when the ensemble size is smaller than the dimension of the model's sub-space, both the EnKF and EnOI may require localisation around each observation in order to eliminate effects of sampling error and to increase the effective number of independent ensemble members used to construct an analysis. However, we find that localisation can degrade an analysis if the length scales of the localising function are too short. We demonstrate that as the length-scale of the localising function is decreased, localisation can significantly compromise the model's dynamical balances. We also find that localisation artificially amplifies high frequencies for applications of the EnKF. Based on our experiments, for applications where localisation is necessary, the length-scales of the localisation should be larger than the decorrelation length-scales of the variables being updated.
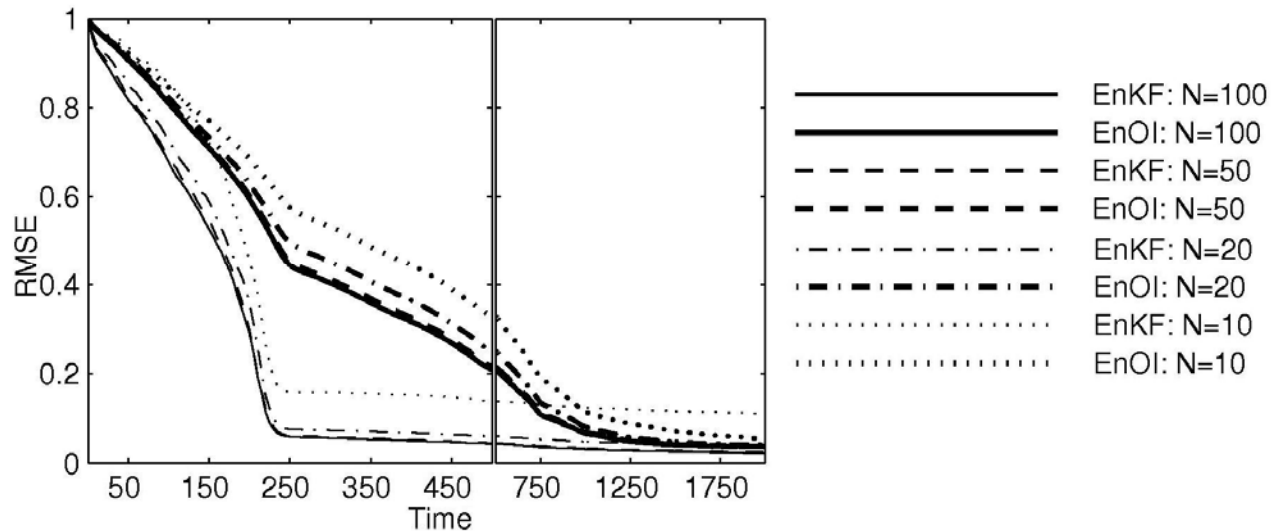

Figure 1: Root-mean-squared error (RMSE) for a linear advection model versus time, for an EnKF and EnOI with different ensemble sizes, N (ranging from 10-100) with covariance localisation; showing that while EnOI is less optimal than EnKF it can produce comparable results (N.B. EnOI costs less than 1/N times that of the EnKF); adapted from Oke et al. (2006).

Comparison of ensemble-based filters:

ESRFs update an ensemble of forecasts in two steps. In the first step, the ensemble mean is updated using the analysis equation from Kalman filter theory. In the second step, the ensemble anomalies (perturbations) are transformed (updated) so that their covariance matches the theoretical analysis error covariance from Kalman filter theory. There are a number of algorithms for this transformation (e.g.,

Tippet et al. 2003; Evensen 2004). The most efficient of these ESRFs is the Ensemble Transform Kalman Filter (ETKF; Bishop et al. 2001). We demonstrate that only a subset of ESRF algorithms that are currently employed preserve the ensemble mean (from step one). As a result, the ensembles covariance no longer matches the theoretical values (in step two), as ESRF theory intends. This renders the filter sub-optimal. We demonstrate the significance of this through a series of experiments with two small models; a linear advection model (Evensen 2004; a simple, stationary, linear model) and the Lorenz-40 model (Lorenz and Emmanuel 1998; a complex, highly non-linear model that is often used to evaluate data assimilation schemes (e.g., Whitaker and Hamill 2002; Ott et al. 2004). No inflation (where one artificially increases the ensemble spread about the mean) was used for the linear advection model; but a range of inflations (from 0-10%) were tested for the Lorenz-40 model. Figure 2 shows the root-mean-squared error (RMSE) versus ensemble size for both models. For the Lorenz-40 model, the results are shown for the inflation that gave the smallest RMSE for a given ensemble size; from Figure 3, where RMSE is plotted as a function of ensemble size and inflation factor. Clearly, the ESRFs that use a mean-preserving transformation (i.e., symmetric and mean-preserving random rotations (RR)) are superior to the other ESRFs and the traditional EnKF. The foot-print of the contours in Figure 3 is an indication of the robustness of each filter (white squares indicate filter divergence); and smaller inflation is generally desirable. The mean-preserving ESRFs are the most robust, most accurate and require the least inflation.

While Wang et al. (2004) reported a "small improvement" using a mean-preserving filter over a non-mean-preserving filter in an atmospheric general circulation model, we suggest that their chosen metric (an energy norm) for assessing the performance of different filters was inappropriate. As a result, we suggest that Wang et al. (2004) understate of the importance of using a mean-preserving transformation. A number of recent studies have used non-mean-preserving solutions (e.g., Evensen 2004; Leeuwenburgh 2005; Leeuwenburgh et al. 2005; Torres et al. 2006), indicating that the importance of preserving the mean during the ensemble transformation is not generally appreciated. Based on both our understanding of the ESRF theory and our numerical experiments, we argue that only mean-preserving solutions should be used in practical data assimilation.
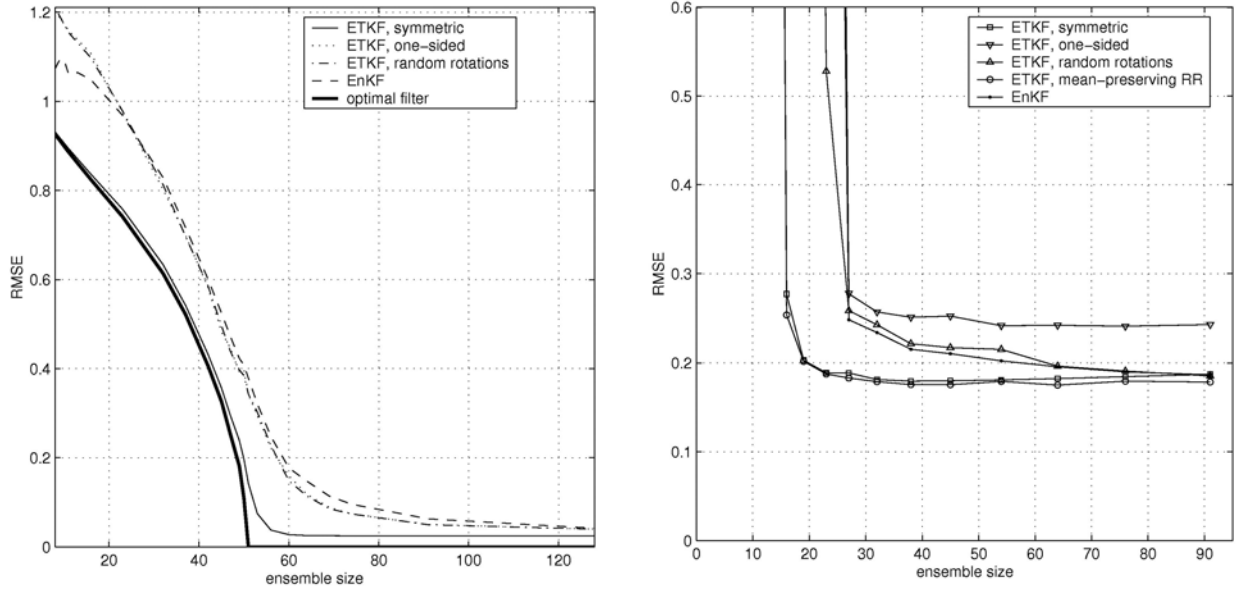
Figure 2: RMSE versus ensemble size for the linear advection model (left; with no inflation; model dimension = 51) and for the Lorenz-40 model (right; for the "best" inflation from Figure 3; model dimension = 27) for different ensemble-based filters (mean preserving filters include ETKF, Symmetric and ETKF, mean-preserving RR). The superior performance of the mean-preserving ESRFs is evident by smaller RMSE for all ensemble sizes; adapted from Sakov et al. (2006).
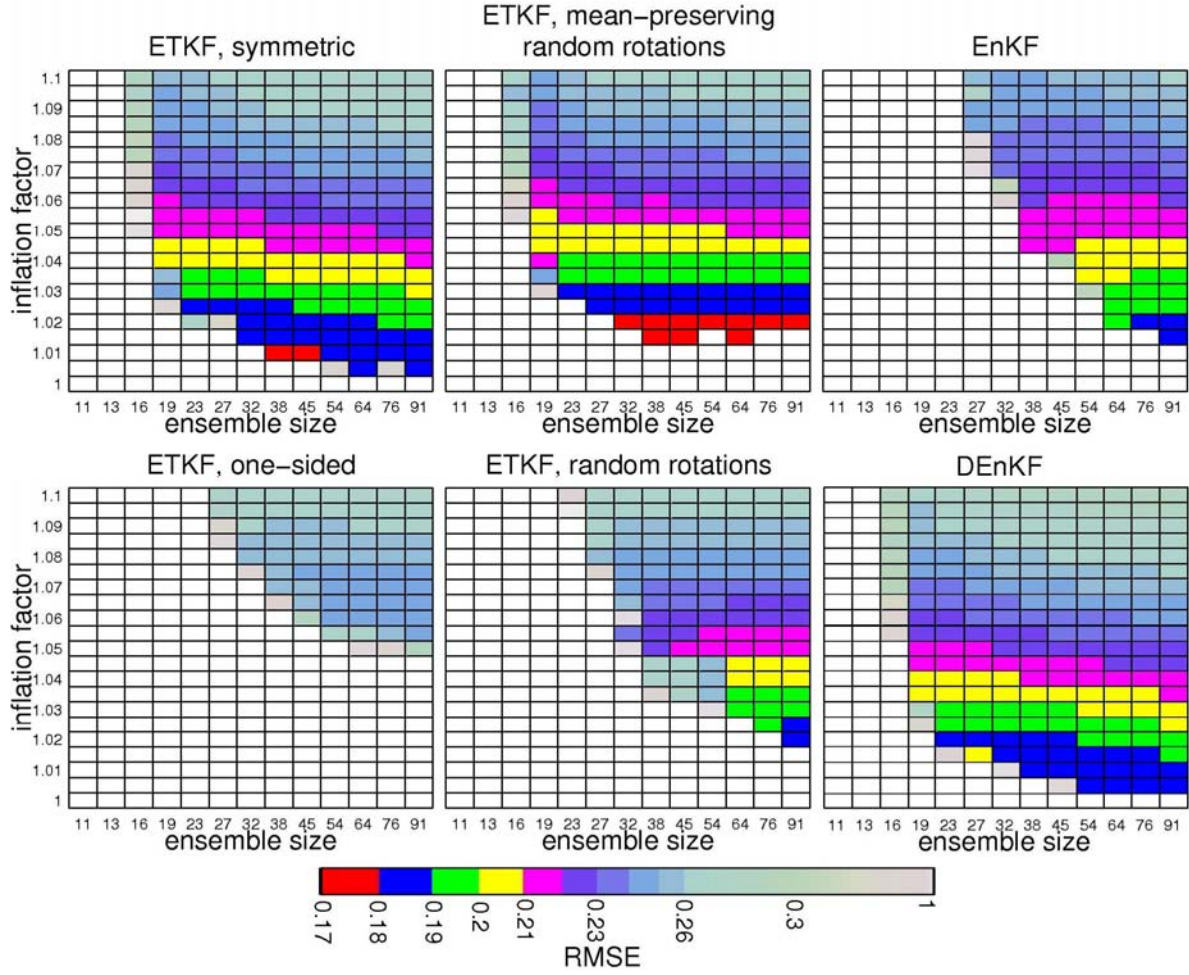
Figure 3: RMSE versus ensemble size and inflation factor for the Lorenz-40 model for different filters showing the superior performance of the mean-preserving filters (ETKF, symmetric and ETKF, mean-preserving random rotations) and the DEnKF; experiments that do not converge and white; adapted from Sakov et al. (2006) and Sakov and Oke (2006).

A deterministic EnKF:

We propose a new formulation for the EnKF that boasts the performance and robustness of the ESRF with the simplicity and versatility of the EnKF. We refer to this filter as the Deterministic EnKF (DEnKF). Following the original proposition of the EnKF (Evensen 1994), Burgers et al. (1998) noted that a straightforward implementation of the EnKF results in a premature ensemble collapse. They proposed the use of perturbed observations as a solution. This has generally become the standard for the EnKF; and is typically employed, along with covariance localisation for practical applications (e.g., Houtekamer and Mitchel 2001). However, the EnKF with perturbed observations can become problematic when a small ensemble is used because of sampling error. This has led some researchers to employ deterministic ESRFs (e.g., Tippett et al. 2003; Evensen 2004) that do not suffer from sampling error and are therefore more robust; particularly when a small ensemble is used. However, we note that the inclusion of covariance localisation with the ESRF requires observations to be assimilated serially, one at a time, which becomes computationally inefficient when many observations

are used. The DEnKF, proposed by Sakov and Oke (2006), boasts a performance that is comparable to the ESRFs (for all cases we've considered), but readily permits localisation, like the EnKF. We've shown that without perturbed observations, a straightforward implementation of the EnKF yields an analysis error covariance that can be expressed as:

$$P^a = P^f - 2KHP^f + KHP^f H^T K^T ,$$ (1)

where P is the error covariance and superscripts a and f denote analysis and forecast respectively; K is the Kalman gain and H is the observation operator. By neglecting the quadratic term in (1), which is valid when K is small, we note that the gain is twice as large as it should be to match the theoretical analysis error covariance, $(I - KH)P^f$. We therefore suggest a simple alternative to the traditional EnKF, where the ensemble mean is updated using the standard analysis equation from Kalman filter, just like an ESRF; and the ensemble anomalies are also updated using the analysis equation from the Kalman filter, but with half the Kalman gain:

$$A^a = A^f - \frac{1}{2} KHA^f ,$$ (2)

where A denotes an ensemble of anomalies. Using this formulation, the ensemble analysis error covariance approximately matches the theoretical values, $(I - KH)P^f + \frac{1}{4} KHP^f H^T K^T$.

A comparison between an ESRF with a mean-preserving transformation (see ETKF, symmetric and ETKF, mean-preserving random rotations in Figure 3) and the DEnKF for the Lorenz-40 model is shown in Figure 3. We find that the DEnKF performs almost as well as the mean-preserving ESRFs; and that it is clearly superior to the EnKF with perturbed observations, particularly for small ensemble sizes.

The DEnKF has many practical advantages over both the ESRF and the traditional EnKF. The DEnKF retains the simplicity and versatility of the traditional EnKF. That is, it can be developed and implemented relatively easily; and it readily permits the use of localisation (Houtekamer and Mitchel 2001; Hamill et al. 2001) like the EnKF. By contrast, the ESRF does not readily permit localisation, as noted above. It can be shown that the DEnKF is a linearization of the ESRF; and is deterministic. Consequently, it does not suffer from the same sampling error that can degrade the performance of the EnKF when small ensemble are used. Compared to the ESRF, the DEnKF will always over-estimate the ensemble spread, rather than under-estimating it. This means that while the filter may not converge as quickly or quite perform as optimally as an ESRF (Figure 3), it will be more conservative, and more "forgiving" of mis-specified errors; and is not as susceptible to ensemble collapse or filter divergence as ESRFs. We argue that for any practical data assimilation system, the DEnKF offers a near-optimal and computationally efficient alternative to the ESRF.

IMPACT/APPLICATIONS

A new version of the EnKF has been proposed, here referred to as the DEnKF. This filter boasts the computational simplicity and efficiency of the traditional EnKF; but performs almost as well as the more complicated ESRFs. The DEnKF has the advantage over the ESRF that it readily permits covariance localisation that is typically required for most practical applications. We argue that the DEnKF should be considered for any application of an ensemble-based assimilation system.

Based on our research, we argue that if an ESRF is to be used; then only mean-preserving transformations should be used in practice. All other ESRFs are sub-optimal.

This research involves the development of a data assimilation tool that will be suitable for a variety of ocean models and configurations. We are particularly focused on making this tool useful for short-range prediction of mesoscale variability. We anticipate that this tool will be valuable for all agencies involved in this project.

RELATED PROJECTS

The BLUElink: Ocean Forecasting Australia project (http://www.marine.csiro.au/bluelink/) is a partnership between the CSIRO Wealth from Oceans Flagship Program, the Australian Bureau of Meteorology and the Royal Australian Navy. An EnOI version of BODAS has been developed and applied to OFAM to perform the Bluelink ReANalysis (BRAN) that spans 1992-2004. BRAN assimilates altimetric and coastal sea-level, and in situ temperature and salinity observations from all available sources (e.g., Argo, WOCE, TAO, SOOP XBT). Comparison with independent observations indicates that BRAN successfully represents the true history of the near-surface mesoscale circulation around Australia for the past decade. For example, Figure 4 demonstrates excellent agreement between the paths taken by un-assimilated satellite-tracked surface drifters and the BRAN sea-level anomaly fields off eastern Australia during 2000. BRAN uses an EnOI version of BODAS with covariance localisation. This experiment indicates that a simple EnOI system can constrain a realistic, eddy-resolving ocean general circulation model (Oke et al. 2005; Schiller et al. 2006). We therefore argue that this approach is suitable for operational ocean prediction.
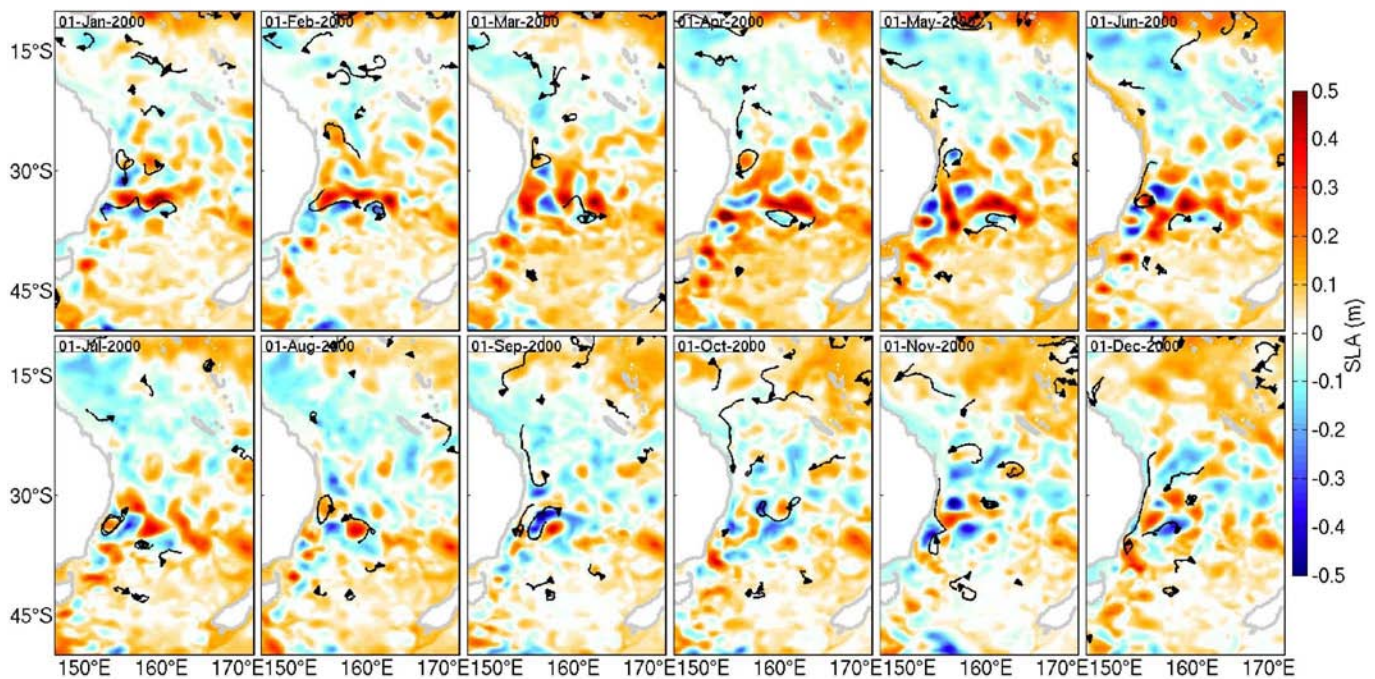
Figure 4: A sequence of reanalysed sea-level anomalies (monthly means) off eastern Australia from BRAN and surface drifter paths for the 1 month period centered at the specified date in 2000. Negative (cyclonic) anomalies are blue; positive (anti-cyclonic) anomalies are red; zero is white; adapted from Oke et al. (2005).

Research on optimal array design and the improved ensemble data assimilation algorithms have also been supported by the CSIRO Wealth from Oceans Flagship Program (http://www.csiro.au/csiro/ channel/pchba,,.html). A key element of this program is the enhancement of our ocean forecasting capabilities. Under this project we have developed a method for designing an optimal observation array using ensemble-based data assimilation theory (Oke and Schiller 2006; Oke and Sakov 2006a); and developed a method for estimating representation error for oceanic observations (Oke and Sakov 2006b). Much of this research has strong synergies with the project that is the subject of this annual report.

REFERENCES

Bishop, C., B. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. part i: theoretical aspects. *Monthly Weather Review*, **129**, 420–436.

Burgers, G., P. J. van Leeuwen and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review,* **126**, 1719-1724.

Evensen, G., 1994: Sequential data assimilation with a non-linear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research,* **99**, 10143-10162.

Evensen, G., 2004: Sampling strategies and square root analysis schemes for the EnKF. *Ocean Dynamics,* 54, 539-560.

Griffes, S. M., R. C. Pacanowski, and A. Rosati, 2004: A technical guide to MOM4, GFDL Ocean Group Technical Report No. 5. NOAA/Geophysical Fluid Dynamics Laboratory, 371 pp.

Hamill, T. M., J. S. Whitaker and C. Snyder, 2001: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Monthly Weather Review,* **129**, 2776-2790.

Houtekamer, P. L. and H. L. Mitchel, 2001: Data assimilation using an ensemble kalman filter technique. *Monthly Weather Review*, **126**, 796–811.

Leeuwenburgh, O., 2005: Assimilation of along-track altimeter data in the tropical pacific region of a global OGCM ensemble. *Quarterly Journal of the Royal Meteorological Society,* **131**, 2455–2472.

Leeuweburgh, O., G. Evensen, and L. Bertino, 2005: The impact of ensemble filter definition on the assimilation of temperature profiles in the tropical pacific. *Quarterly Journal of the Royal Meteorological Society,* **131**, 3291-3300.

Lorenz, E. N. and K. A. Emanuel 1998: Optimal sites for supplementary weather observations: simulation with a small model. *Journal of Atmospheric Sciences,* **55**, 399-414.

Oke, P. R., J. S. Allen, R. N. Miller, G. D. Egbert and P. M. Kosro, 2002: Assimilation of surface velocity data into a primitive equation coastal ocean model. *Journal of Geophysical Research,* **107**, doi:10.1029/200JC000511.

Ott, E., B. R. Hunt, I. Szunyogh, A. V. Zimin, E. J. Kostelich, M. Corazza, E. Kalnay, D. J. Patil, and J. A. Yorke, 2004: A local ensemble Kalman filter for atmospheric data assimilation. *Tellus*, **56A**, 415–428.

Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill and J. S. Whitaker, 2003: Ensemble square root filters. *Monthly Weather Review,* **131**, 1485-1490.

Torres, R., J. I. Allen, and F. Figueiras, 2006: Sequential data assimilation in an upwelling influenced estuary. *Journal of Marine Systems,* **60**, 317–329.

Wang, X., C. H. Bishop, and S. J. Julier, 2004: Which is better, an ensemble of positive-negative pairs or a centered spherical simplex ensemble? *Monthly Weather Review*, **132**, 1590–1605.

Whitaker, J. S. and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Monthly Weather Review*, **130**, 1913–1924.

PUBLICATIONS

Oke, P. R., A. Schiller, D. A. Griffin and G. B. Brassington, 2005: Ensemble data assimilation for an eddy-resolving ocean model, *Quarterly Journal of the Royal Meteorology Society,* **131**, 3301-3311.

Oke, P. R., P. Sakov, S. P. Corney, 2006: Impacts of localisation in the EnKF and EnOI: Experiments with a small model, Ocean Dynamics, http://dx.doi.org/10.1007/s10236-006-0088-8.

Oke, P. R., and A. Schiller, 2006: A model-based assessment and design of a tropical Indian Ocean mooring array, *Journal of Climate,* [in press, refereed].

Oke, P. R., and P. Sakov, 2006a: Optimal array design: Applications to the tropical Indian Ocean. Monthly Weather Review, [submitted, refereed].

Oke, P. R., and P. Sakov, 2006b: A method for estimating representation error of oceanic observations. Geophysical Research Letters, [submitted, refereed].

Sakov, P., P. R. Oke and S. P. Corney, 2006: Only mean-preserving transformations should be used in the ensemble square root filter. *Monthly Weather Review,* [submitted, refereed].

Sakov, P., and P. R. Oke, 2006a: A deterministic ensemble Kalman filter. *Monthly Weather Review, [*submitted, refereed].

Schiller, A., P. R. Oke, G. Brassington, M. Entel, R. Fiedler, D. A. Griffin, J. V. Mansbridge, G. A. Meyers, K. R. Ridgway, and N. R. Smith, 2006: Eddy-resolving ocean circulation in the Asian-Australian region inferred from an ocean reanalysis effort. *Progress in Oceanography,* [submitted, refereed].

Preprints of submitted manuscripts are available from http://www.cmar.csiro.au/staff/oke/.